VLSI IMPLEMENTATION OF A PROGRAMMABLE CURRENT-MODE NEURAL NETWORK

YASUHIRO OTA and BOGDAN M. WILAMOWSKI
Dept. of Electrical Engineering, University of Wyoming, Laramie, WY, USA

ABSTRACT:
In analog signal processing, knowledge is encapsulated in the form of analog synaptic weights. In simple hardware systems, these weights are determined by resistance values. The analog computation of a scalar product and the subsequent nonlinear mapping can be performed by a summing amplifier with saturation. In this paper, a design for a general-purpose programmable, current-mode neural network using VLSI circuits operated in a subthreshold mode is proposed. The proposed neuron circuit consists of a CMOS differential pair, and the synapse circuit uses current mirrors controlled by static RAM (SRAM). A 2-link robot forward kinematics has been implemented in our neural chip design.

INTRODUCTION

A significant research activity has been developed around artificial neural networks in the past two decades. From a practical point of view, VLSI implementation of neural networks is very important because neural networks simulated on a digital computer are not able to take an advantage of parallel processing. VLSI technology allows us not only to simulate biological systems, but also to emulate them in designing artificial sensory systems in real-time. Many approaches on artificial neural network implementations in VLSI technology have been reported in the literature (Watanabe *et al.*, 1993; Graf and Jackel, 1989; Atlas and Suzuki, 1989; Morishita *et al.*, 1990; Boahen *et al.*, 1989). Bipolar transistors were used for reducing multiplier error based on a bipolar-MOS analog circuit (Morishita *et al.*, 1990). In our design, the MOS transistors are operated under subthreshold region to keep the power consumption low. An on-chip SRAM cell array is used to digitally store 8-bit synaptic weights in this paper. The digital memory is the most straightforward implementation of a quantized analog storage. Another advantage to digital circuits is that they provide for the high arithmetic precision requirements of many existing learning algorithms. In general, 8 and sometimes 16 bits of precision are necessary for the synaptic weight representation for hardware implementation. Wawrzynek *et al.* (1993) also showed that 16-bit weight values and 8-bit output values were sufficient to achieve training and classification results comparable to 32-bit floating-point operation . The digital storage using an SRAM provides several advantages over using analog storage on an SRAM. The first advantage is the easy programmability. The analog storage requires a highly sensitive voltage sensor circuitry since it divides the voltage of the cell node into 2^n levels to store n-bit data.

In this paper, VLSI architecture for a subthreshold operation of a programmable general-purpose neural network is proposed. For a neuron (processing element) circuit, a CMOS differential pair is utilized. For a synapse circuit, current multipliers are utilized to implement digital binary weights. An on-chip SRAM cell array is included so that the weights can be programmed.

The remainder of this paper is organized as follows. The next section provides the structure of the subthreshold CMOS neuron and synapse circuits in detail; including the operation of the network with simulation results of the neuron circuit. Then the following section illustrates an implementation of a 2-link robot forward kinematics as a case study of this network design and its experimental results. The final section concludes this paper with some possible future issues that need to be addressed.

STRUCTURE OF A PROGRAMMABLE NEURAL NETWORK

A block diagram for hardware implementation of the current-mode neural network is shown in Figure 1. This diagram shows a case with a single neuron network; however, one can construct and extend to any single- or multi-layered network by conserving this concept. The neuron circuit consists a differential pair with Wilson current mirrors to obtain current-mode, positive and negative activation functions as shown in Figure 2(a). A simple and unique current-to-voltage converter has been designed and utilized in the neuron circuit to convert the sum of currents from the synapse circuit cells in the previous layer (simply the input current for the input stage) to voltage that is applied to the gate of input node of the differential pair as shown in Figure 2(b). This current-to-voltage converter is also designed such that it avoids polysilicon resistors that would consume relatively large chip area. The reference voltage on the other transistor in the differential pair is set to 2.5V with the voltage source circuit shown in Figure 2(c). Notice that this is simply a voltage divider using an *n*-channel transistor and a *p*-channel transistor as resistors, and one can obtain any desired voltage-division ratio by choosing proper *W/L* ratios of the MOS transistors. The transfer characteristics of the unipolar sigmoid function is illustrated in Figure 3. In this design, both positive and negative sigmoid functions can be available; however, only the positive sigmoid function is used for a unipolar neuron.

Figure 4 shows the circuit schematic of an 8-bit digital programmable synapse cell with a sign bit. The transistors M1-M8 are sized such that the *W/L* ratios are 2/2, 4/2, 8/2, 16/2, 32/2, 64/2, 128/2, and 256/2 (μm) to obtain an 8-bit binary multiplication. These transistors can be programmed via an SRAM cell array so that one can obtain any 8-bit weights. The sign-bit controls the direction of the output current and is turned on for a negative current output. The transistors M10-M14 have the same size; therefore, when the sign-bit is turned on, the drain current of M10 is copied to that of M13 and M14. By pulling down the drain current of M11 with twice as much using M13 and M14, the direction of the output current becomes opposite compared to a positive output current. Consequently, a negative current or a negative weight can be produced. In Figure 4 the direction of negative current flow is to the left. Notice that typical sigmoidal activation functions are

$$f(net) \propto \lambda net \tag{1}$$

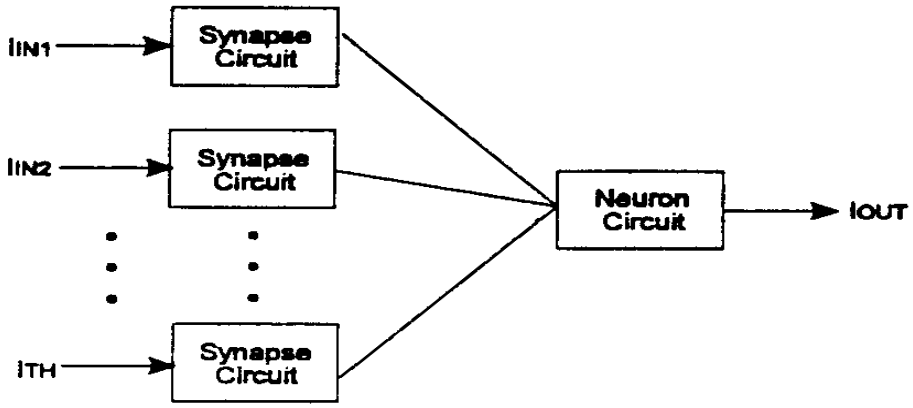

Figure 1: Block diagram of the current-mode neural network VLSI implementation.

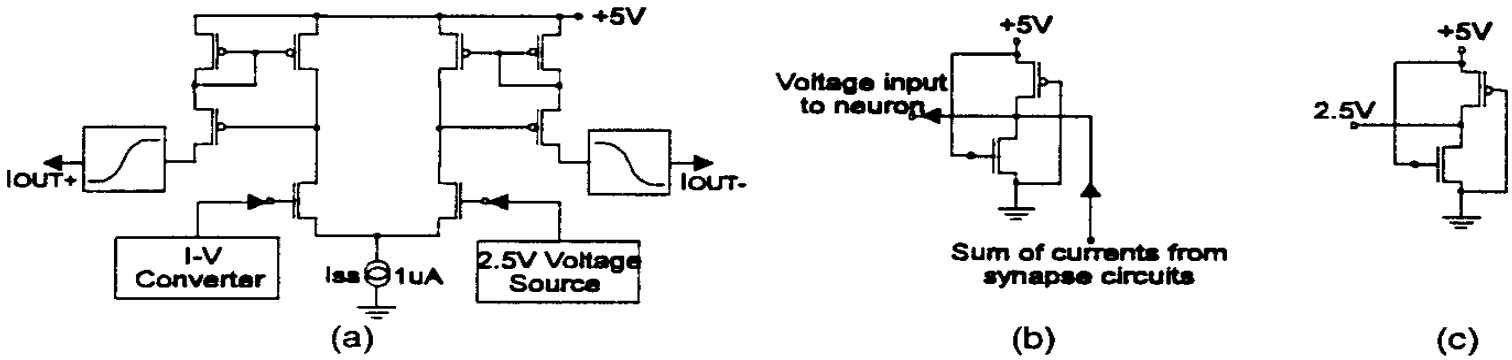

Figure 2: Neuron circuit using MOS differential pair with Wilson current mirrors: (a) Neuron circuit, (b) Current-to-voltage converter, and (c) 2.5V voltage source. All the transistors have *W/L* ratio of $2\mu m/2\mu m$.

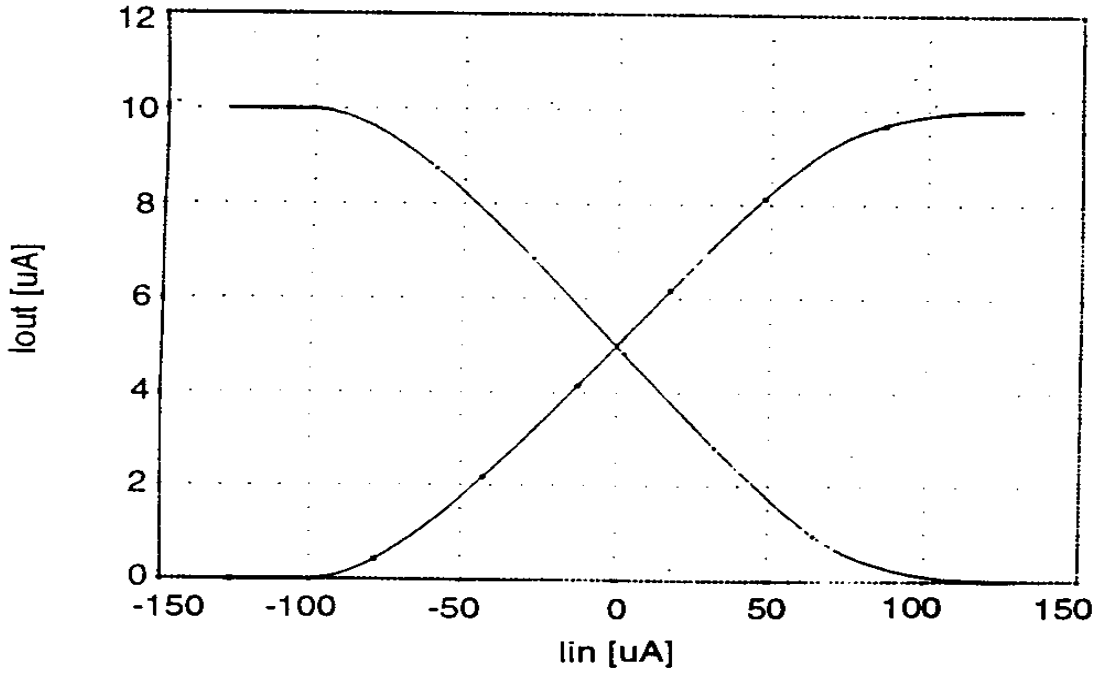

Figure 3: Transfer characteristics of the unipolar neuron circuit.

where $\lambda > 0$ and is proportional to the neuron gain determining the steepness of the continuous function $f(net)$ near $net = 0$. Thus, instead of controlling the value of λ in the sigmoidal activation function, it is possible to compensate the right-hand term of (1) by controlling the *net* value or by multiplying a certain fraction to the weights. This compensation can be accomplished by adjusting the *W/L* ratios of transistors M9 and M10 in Figure 4.

In some neural network applications, such as forward and/or inverse robot kinematics, the weights are previously known, and such a system needs not be flexible to adjust weights. In such a case, the synapse circuit diagram shown in Figure 4 can be significantly simplified. Instead of using the circuit in Figure 4 for a programmable weight, the circuit shown in Figure 5 with a fixed weight by adjusting a proper W/L ratio of the transistors, Mref and M1, can be used. If a known weight is positive, then the *p*-channel current mirror (M2 and M3) is necessary; otherwise only the *n*-channel current mirror (Mref and M1) will suffice.

SIMULATION AND LAYOUT DESIGN

As a case study of this neural chip, a simple 2-link robot forward kinematics has been tested, and the functionality of the circuit has been verified with the SPICE circuit simulator. The layout design is based on a standard 2-μm CMOS technology.

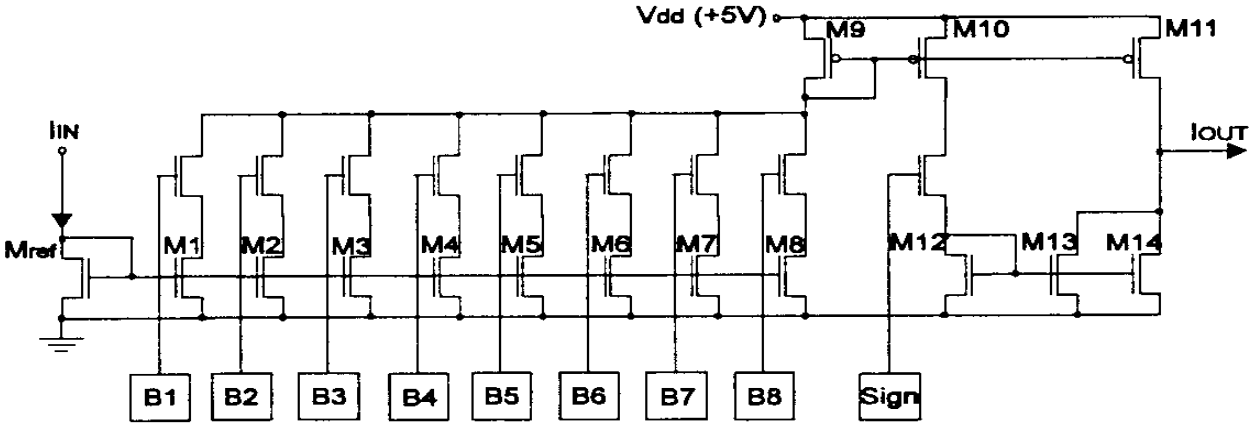

Figure 4: 8-bit digital programmable synapse circuit with a sign bit. Transistors M1-M8 are sized as 2/2, 4/2, 8/2, 16/2, 32/2, 64/2, 128/2, 256/2. All other transistors are of minimum size. B1-B8 and the sign bit can be programmed and controlled by an SRAM cell array.

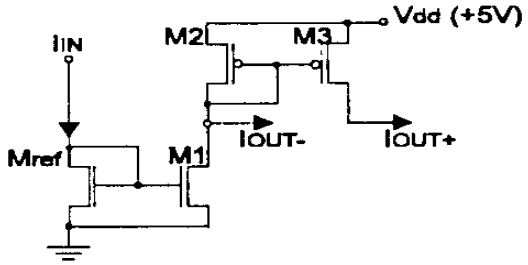

Figure 5: Fixed-weight synapse cell.

Figure 6 shows the 2-link manipulator used in this design. Here a point mass is assumed for simplicity. Craig (1989) gives an excellent explanation of detail robot kinematics. For this particular 2-input/2-output forward kinematics design, there are three hidden neurons and two output neurons, and thus the number of synapse circuits required is seventeen (nine for the hidden layer and eight for the output layer) as shown in Figure 7. Some of the experimental results are listed in Table 1. As can be seen from this table, the results using our proposed design are promising compared to the desired outputs. The error is caused mainly due to the training error which is 0.0787 in this example.

Since the synaptic weights can be programmed via an SRAM cell array, some other applications that have the same network architecture as this example can be implemented as well. The entire area required for the layout design is approximately $3200\mu m \times 3200\mu m$.

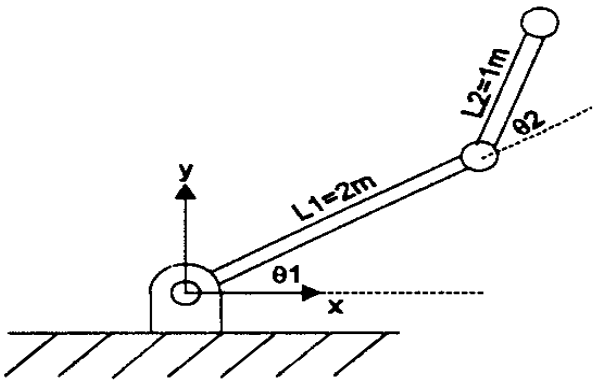

Figure 6: 2-link planar robot manipulator. Point mass is assumed for simplicity.

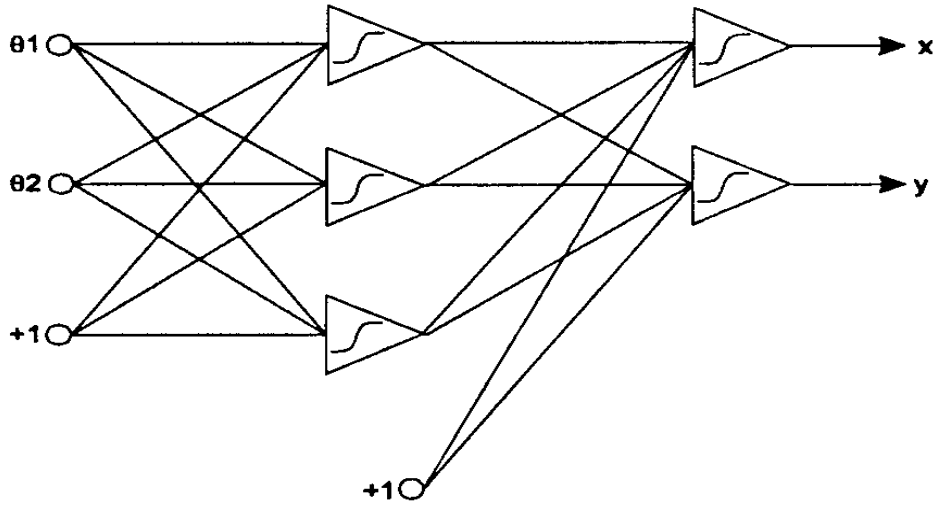

Figure 7: Two-layer neural network structure for the 2-link robot forward kinematics example.

Table 1. Circuit simulation results of some experimental forward kinematics computation. (Average sum-of-square error of the network training in this experiment is 0.0787.)

Experiment	In put		Out put					
	Joint Angles		Desired Position		Position w/ Actual Weights		Position w/ Proposed Design	
	θ1[deg]	θ2[deg]	x [m]	y [m]	x [m]	y [m]	x [m]	y [m]
1	60	340	1.7667	2.3748	1.7275	2.3331	1.7454	2.4039
2	320	325	1.7909	-2.2516	1.7993	-2.0052	1.7787	-2.0521
3	0	45	2.7072	0.7072	2.6673	0.7041	2.6616	0.7199
4	240	305	-1.9962	-1.8192	-1.9152	-1.9464	-2.1418	-1.8444

CONCLUSIONS

VLSI architecture for a subthreshold operation of a programmable, general-purpose neural network has been discussed in this paper. Virtually, only nine MOS transistors are necessary to produce a positive sigmoid function in this proposed design. As described in the previous section, the experimental results of the forward kinematics example using the proposed circuit structure are promising, and such a practical system can be implemented in real-time.

The approach that has been described in this paper is not the sole solution for realizing a fairly large, practically useful artificially neural network in the near future. It may be that the total number of neurons and synapses in the network should not be limited to certain values. In case of large network systems, a wafer-scale integration, or other digital approach, such as using DSP boards, may be powerful and useful.

REFERENCES

Atlas, L. E., Suzuki, Y., (1989). Digital Systems for Artificial Neural Networks, *IEEE Circuits and Devices Mag.*, Nov, 20-24.

Boahen, K. A., Pouliquen, P. O., Andreou, A. G., and Jenkins, R. E. (1989). A Heteroassociative Memory Using Current-Mode MOS Analog VLSI Circuits, *IEEE Trans. on Circuits and Systems*, 36 (5), May, 747-755.

Craig, J. J., (1989). *Introduction to Robotics: Mechanics and Control*, 2nd ed., Addison-Wesley: Reading, MA.

Geiger, R. L., Allen, P. E., and Strader, N. R. (1990). *VLSI Design Techniques for Analog and Digital Circuits*, McGraw-Hill: New York.

Graf, H. P., Jackel, L. D., (1989). Analog Electronic Neural Network Circuits, *IEEE Circuits and Devices Mag.*, July, 44-55.

Morishita, T., Tamura, Y., and Otsuki, T. (1990). A BiCMOS Analog Neural Network with Dynamically Updated Weights, *IEEE Int. Solid-State Circuits Conf. Digest of Tech. Papers*, Feb, 142-143.

Simpson, P. K., (1992). Foundations of Neural Networks, *Artificial Neural Networks: Paradigms, Applications, and Hardware Implementations*, E. Sanchez-Sinencio and C. Lau, Eds., IEEE Press: New York.

Watanabe, T., Kimura, K., Aoki, M., Sakata, T., and Ito, K. (1993). A Single 1.5-V Digital Chip for a 10^6 Synapse Neural Network, *IEEE Trans. on Neural Networks*, 4 (3), May, 387-393.

Wawrzynek, J., Asanovic, K., and Morgan, N. (1993). The Design of a Neuro-Microprocessor, *IEEE Trans. on Neural Networks*, 4 (3), May, 394-399.

Zurada, J. M., (1992). *Introduction to Artificial Neural Systems*, West: St. Paul, MN.

INTELLIGENT ENGINEERING SYSTEMS THROUGH ARTIFICIAL NEURAL NETWORKS

VOLUME 4

Editors:

Cihan H. Dagli
Benito R. Fernández
Joydeep Ghosh
R. T. Soundar Kumara